Letters

Nonsymmetric PDF Estimation by Artificial Neurons: Application to Statistical Characterization of Reinforced Composites

Simone Fiori

Abstract—**We present a generalized adaptive activation function neuron structure which learns through an information-theoretic-based principle, which is able to estimate the probability density function of incoming input. It provides a low-order smooth robust estimate of the input signal probability density function. The presented method has been developed with reference to statistical characterization of polypropylene composites reinforced with vegetal fibers, that the proposed numerical experiments pertain to.**

***Index Terms*—Artificial neurons, probability density function (PDF) estimation, reinforced polypropylene composites.**

I. Introduction

Probability density function (PDF) modeling from incomplete data may be regarded as a constrained functional approximation problem [8]. It concerns the estimation of the PDF of a signal when some particular features of the true PDF are observed (measured) or obtained from signal's samples.

Generally speaking, the problem of modeling a probability density function, given a finite number of data points, plays a central role in statistical signal processing, with applications ranging from econometrics [5] to industrial engineering [6], [7]. Several techniques were developed in order to tackle this problem, which may be classified mainly in parametric, nonparametric, and semiparametric techniques [4].

In particular, semiparametric methods are not restricted to specific functional forms and, also, the size of the model only grows with the complexity of data-space structure, not simply with the number of available data. As two semiparametric models, it is worth citing the mixture-of-kernel and the multilayer perceptron (MLP) with learnable parameters [4]. The mixture model consists of a linear adjustable superposition of simple flexible kernels; both kernels' parameters and combination factors may be adjusted by means of maximum-likelihood-based procedures, as the well-known expectation-maximization (EM) algorithm [10].

The aim of this contribution is to present a generalized adaptive activation function neuron (FAN) structure and illustrate its behavior through numerical experiments performed on real-world data, with particular emphasis to statistical characterization of propylene composites reinforced with vegetal fibers.

II. PDF Modeling by FAN Neurons

Let $x(t)$ be a stationary random process with smooth unknown probability density function $f_x(x)$. The problem of estimating $f_x(x)$ is closely related to the estimation of the cumulative distribution function $F_x(x)$ and of the "score function" $s_x(x) \stackrel{\text{def}}{=} f'_x(x)/f_x(x)$.

In principle, estimating $f_x(x)$, $F_x(x)$ or $s_x(x)$ is totally equivalent, however, $F_x(x)$ is the only one among these which is a bounded saturating (sigmoidal) function, thus it appears to be naturally estimable by neurons that provide saturating input–output transferences. Moreover, a way of estimating $f_x(x)$ relies on the "flattening" property of the function $F_x(x)$: It is, in fact, well known [20], [22] that warping the process $x(t)$ by the function $F_x(x)$, that is computing $y(t) = F_x(x(t)) \in [0, 1]$, results in a uniformly distributed process or, equivalently, a process with maximum entropy H_y. Thus when $F_x(x)$ is approximated by a monotonic parametric function $y = \mathcal{M}(x; \bar{p})$ with parameters in vector $\bar{p}$, a maximum entropy estimator is obtained by recursively maximizing H_y with respect to $\bar{p}$. The maximum entropy estimator is completely equivalent to the maximum likelihood estimator [23] in the case that there are no prior beliefs biasing the solution, and to Kullback–Leibler estimator [1].

Manuscript received October 22, 2002; revised December 23, 2002.
The author is with the Faculty of Engineering, Perugia University, I-05100 Terni, Italy (e-mail: sfr@unipg.it).
Digital Object Identifier 10.1109/TNN.2003.813825

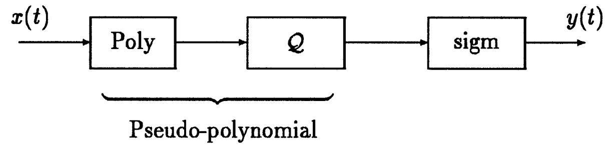

Fig. 1. Abstract structure of the proposed artificial neuronal model.

If $x(t)$ is supposed to be a stationary random process then y will be a random process $y = y(t)$ too, with a PDF denoted here with $f_y(y; \bar{p})$, because it depends upon the value of the parameters. The differential entropy of the random process $y(t)$, defined as

$$H_y(\bar{p}) \stackrel{\text{def}}{=} -\int_0^1 f_y(\zeta; \bar{p}) \log f_y(\zeta; \bar{p}) d\zeta \tag{1}$$

can be related to the differential entropy of the random process $x(t)$ by means of the fundamental formula $f_y = f_x/|\psi|$, where $\psi(x; \bar{p}) \stackrel{\text{def}}{=} \mathcal{M}'(x; \bar{p})$. Using that substitution in the entropy formula (1), yields

$$H_y = -\int_{-\infty}^{+\infty} \frac{f_x}{|\psi|} \log\left(\frac{f_x}{|\psi|}\right) |\psi| d\xi = H_x + \int_{-\infty}^{+\infty} f_x \log |\psi| d\xi.$$

These formulas hold by the hypothesis of monotonicity, which ensures function ψ nowhere changes its sign.

As mentioned, our aim is to maximize the entropy of the neuron's response, which is equivalent to maximize the quantity $G_H \stackrel{\text{def}}{=} H_y - H_x$, referred to as entropy gap between the input and the output random processes.

A. Nonlinear Flexible Neuron Models

Known examples of nonlinear flexible neuron models are given by Pao's functional-link networks and the tensor-product networks (for a detailed review see [11]), the Fourier neural networks, described by Dastani [9] and recently revised by Silvescu [21], based on learnable combinations of sinusoidal basis functions, and the wavelet-based models [21], [24]. A recent review of some nonlinear adaptive activation function neuron models may be also found in [15]. Moreover, several models of nonlinear neurons with filtering synapses have been proposed, which allows taking into account the dependence of the behavior of biological neurons from the temporal structure of the excitations. Among the many models and engineering applications available in the scientific literature, a pervasive example finds e.g., in [3].

However, the mentioned models are almost exclusively restricted to supervised settings. It is, therefore, of interest to extend their learning theories to unsupervised (information-theoretic) tasks.

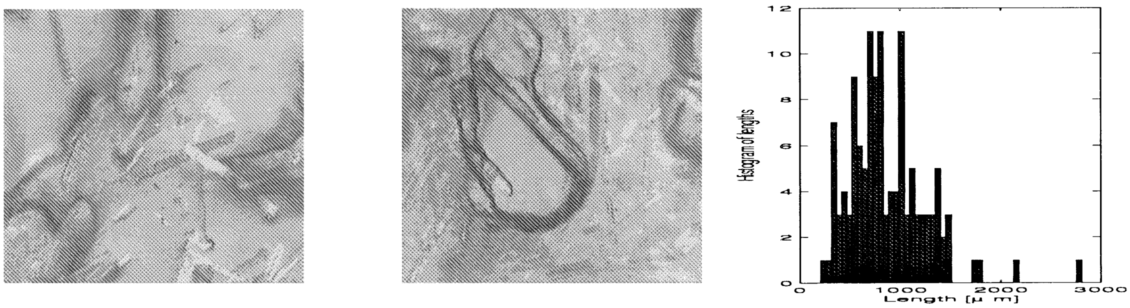

Fig. 2. Two views of a polypropylene matrix with some flax fibers and a histogram-based estimation of lengths distribution.

B. Structural/Learning FAN Equations

In the theoretical seminal paper [12], and later on in the following research works [14]–[18], a particular input–output description for FANs was assumed

$$y = \mathcal{M}(x; \bar{p}) = \mathrm{sigm}[\pi(x; \bar{p})] \tag{2}$$

where $\mathrm{sigm}(\cdot) \in [0, 1]$ is a sigmoidal function, bounded above and below, continuous and strictly increasing; $\pi(\cdot)$ is a monotonically nondecreasing pseudopolynomial depending upon $\bar{N}$ parameters in $\bar{p}$.

This representation has the following advantages: 1) the function $\mathcal{M}(x; \bar{p})$ always ranges in [0, 1] and is always monotonically nondecreasing, thus it represents a valid cumulative function and 2) the computation of the optimal value of $\bar{p}$ may be carried out as an unconstrained maximization problem, because the function $G_H(\bar{p})$ is always limited from above by $-H_x$.

The choice of the internal structure of the neuron, represented by the pseudopolynomial $\pi(\cdot)$ is a key point in its design. As a matter of fact, in the selection of the function $\pi(\cdot)$ it is necessary to take into account the desired monotonicity of neuron's transfer function $\mathcal{M}(\cdot)$, which implies $\pi'(\cdot)$ to be always nonnegative. In the mentioned previous works, this result was obtained by selecting $\pi'(\cdot)$ as a polynomial possessing even-power monomial terms only with positive coefficients. This choice has two drawbacks. First, it limits the flexibility of the polynomial because of the cutting of degrees of freedom. Second, unfortunately it makes the $\psi(\cdot)$ an even function which prevents the neuron from being able to estimate asymmetric PDFs. In the following section we propose a more general model by changing the structure of $\pi(\cdot)$ which also generalizes the asymmetric FAN model proposed in [13].

III. Asymmetric FAN Neuron Model

In order to make the squashed pseudopolynomial structure be able to learn general PDFs, we propose the following FAN model:

$$y = \mathcal{M}(x; b, p) \stackrel{\mathrm{def}}{=} \mathrm{sigm}[\pi(x; b, p)] \tag{3}$$

$$\pi(x; b, p) \stackrel{\mathrm{def}}{=} b + \int_0^x \mathcal{Q}\left(\sum_{n=0}^{N} p_n \xi^n\right) d\xi \tag{4}$$

where $\mathrm{sigm}(\cdot)$ denotes again a positive sigmoidal function, and $\mathcal{Q}(\cdot)$ is whatever positive integrable function that is U-shaped for small values of the argument and which saturates for large values of the argument. The learnable parameters are b and p and amount to $\bar{N} = (N+1)+1 = N+2$. Such structure provides a valid PDF estimation.

The proposed neuronal structure may be related to the taxonomy of neural functions surveyed in [11]. In particular, given the abstract structure illustrated in Fig. 1, it is possible to classify the proposed structure as a sigmoidal-output, pseudopolynomial, delocalized activation function, dot-product neuron (with a single input), which is closely related to the already mentioned functional-link model.

Another interesting question is the generalization of the proposed model to multivariate PDF estimation. From a methodological point of view, such extension can be envisaged if the polynomial in the single variable x is thought to as a possible polynomial of many variables $x_1, x_2, \ldots$ appearing with different powers: This simple generalization would allow the neuron to be able to estimate multivariate PDFs. The price of this generalization, like in other methods, would be the consistent growth of the number of free parameters to adapt. This problem falls, however, outside the scope of the present contribution.

To end with the discussion of the proposed structure, it is worth exposing some considerations which led to its definition with a perspective to the final application. Within this letter, we are interested in probability density function estimation in presence of few data (measures) which prove to be not completely reliable. This means that some lack of values and some concentration of values in the raw distribution might be the effect of measurement errors and not the indicators of real data complexity. An estimation technique is sought for which provides a smooth estimate that can be made robust by some parameter tuning. This arguments justify the reason for which low-order ($N = 2, 3$) polynomial approximations are employed in the experiments and the usefulness of the function $\mathcal{Q}$.

About the learning phase, the proposed adaptive activation function neuron may learn through a batch-type gradient-based optimization of the entropy gap with respect to the polynomial's parameter b and p, namely

$$\Delta p_r \propto \left\langle \frac{1}{\psi}\frac{\partial \psi}{\partial p_r} \right\rangle = \left\langle \frac{\mathrm{sigm}''[\pi]}{\mathrm{sigm}'[\pi]}\frac{\partial \pi}{\partial p_r} + \frac{1}{\pi'}\frac{\partial \pi'}{\partial p_r} \right\rangle,\ r = 0, \ldots, N$$

$$\Delta b \propto \left\langle \frac{1}{\psi}\frac{\partial \psi}{\partial b} \right\rangle = \left\langle \frac{\mathrm{sigm}''[\pi]}{\mathrm{sigm}'[\pi]} \right\rangle$$

where $\langle \cdot \rangle$ denotes enseble average and, by the definition

$$\frac{\partial \pi}{\partial p_r} = \int_0^x \mathcal{Q}'\left(\sum_{n=0}^{N} p_n \xi^n\right) \xi^r d\xi,\ \frac{\partial \pi'}{\partial p_r} = \mathcal{Q}'\left(\sum_{n=0}^{N} p_n x^n\right) x^r.$$

The above learning rules imply a learning step-size which may vary with time in order to stabilize their dynamics.

As a case-study, we consider the choice $\mathrm{sigm}(u) = 0.5 + 0.5\mathrm{erf}(u)$ and $\mathcal{Q}(u) = 1 - \exp(-u^2/2\alpha^2)$. The first choice is motivated by the fact that it simplifies the structure of the ratio $\mathrm{sigm}''[\pi]/\mathrm{sigm}'[\pi]$; the shape of $\mathcal{Q}(u)$ provides a way to make the learning process robust against outliers by a proper selection of the parameter α; in fact, by hypothesis it is nearly quadratic for small values of the polynomial and is nearly constant for large values of the polynomial, which normally appear for samples that belong to the tails of the distribution.

After learning, the function $\psi(x; b, p)$ represents the desired estimate of the actual PDF of incoming signal.

It is worth reporting the expressions of function ψ and quantity G_H that, with the mentioned structural choices, read

$$\psi(x; p, b) \propto \exp\left(-\pi^2(x; p, b)\right) \pi'(x; p, b) \tag{5}$$

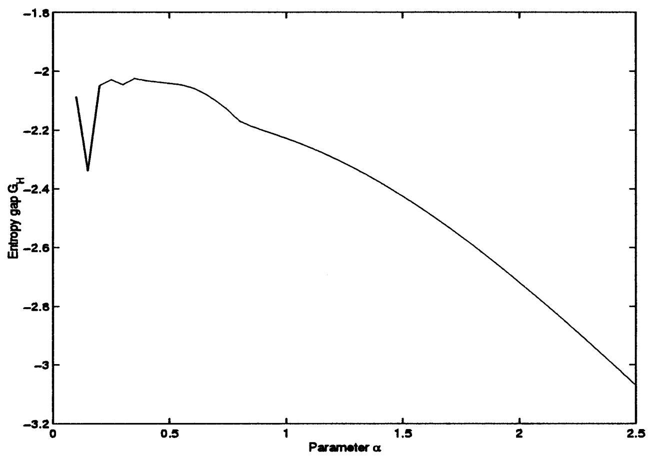

Fig. 3. Result of validation step of structural parameter α.

$$G_H(p, b) = \int_{-\infty}^{+\infty} g_H(p, b|x) f_x(x) dx \tag{6}$$

$$g_H(p, b|x) \propto -\pi^2(x; p, b) + \log \pi'(x; p, b). \tag{7}$$

On a data-point basis, we have $G_H(p, b) \sim \langle g_H(p, b|x) \rangle$.

IV. Experiments on Real-World Data

In order to assess the PDF approximation ability of new asymmetric adaptive activation function neuron, we tested the proposed algorithm on real-world data that exhibit asymmetric distribution. The data-set comes from measures of the length of flax fibers inside slabs of polypropylene composites with 20% concentration of vegetal fibers.

In the technology of polymer composite materials, the vegetal fibers are added to the polypropylene to reinforce its structure: The fibers have initially constant length and diameter, but after the process of transformation of the fibers and the propylene into the yield composite, they spread out and find with different length and diameter values [19]. The micro- and macro-mechanical features of the composite slab, as the elasticity and resistance, depend on the statistical distribution of these parameters and can, e.g., be evaluated through the Halpin–Tsai equation [2], [19].

Two views of a typical matrix of such composite are depicted in Fig. 2, where some randomly oriented discontinuous flax fibers are present. In the first picture, however, the vegetal fibers are well recognizable, while in the second picture it is harder to discern the different fibers. Currently, the measures of lengths and diameters are taken manually from each picture: This explains the noncomplete reliability of the obtained measures. Fig. 2 also shows a histogram-based estimation of the lengths distribution, which shows the data-distribution is asymmetric and quite irregular. The production process usually tends to severely break the fibers, thus the PDF of fibers' length values is expected to be asymmetric [2], with the mean value near the lower values of the distribution; the irregularity is judged as due to the scarcity of data as well as to the measurement errors: No technological meaning is attributed e.g., to the valley between the two large peaks centered in 800 and 1100 μm.

The available 126 different measures come from 20 different pictures of a single polypropylene slab. Prior to entering the PDF-estimation algorithm, the data have been properly normalized (centered and scaled).

The results presented in the following have been obtained by setting $N = 2$ in the FAN neuron in order to obtain a low-order smooth estimation; also, the 126 iterations pertaining to the presentation of any single sample to the neuron have been organized in epochs and the

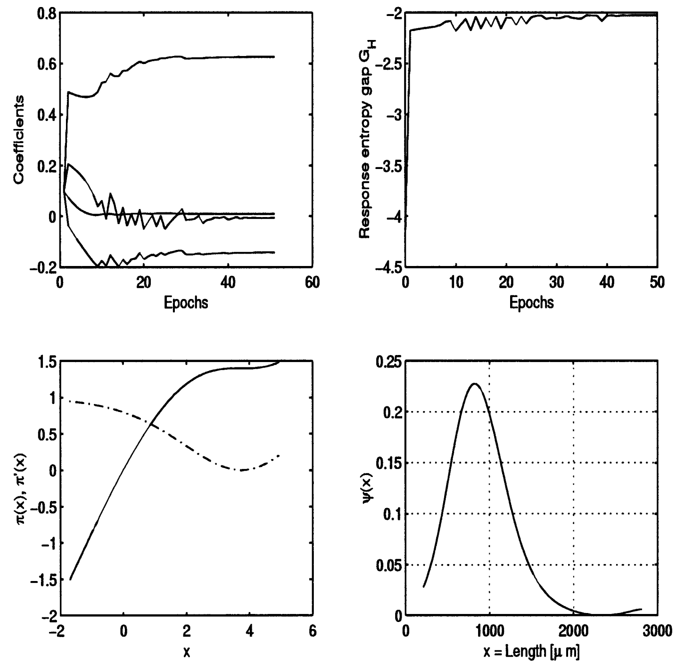

Fig. 4. PDF approximation of lengths of flax fibers embedded in a polypropylene slab for $\alpha = 0.35$. (Bottom-right picture: Arbitrary horizontal scale.) In the bottom-left picture, the solid line refers to the polynomial π while the dot-dashed line refers to its first derivative.

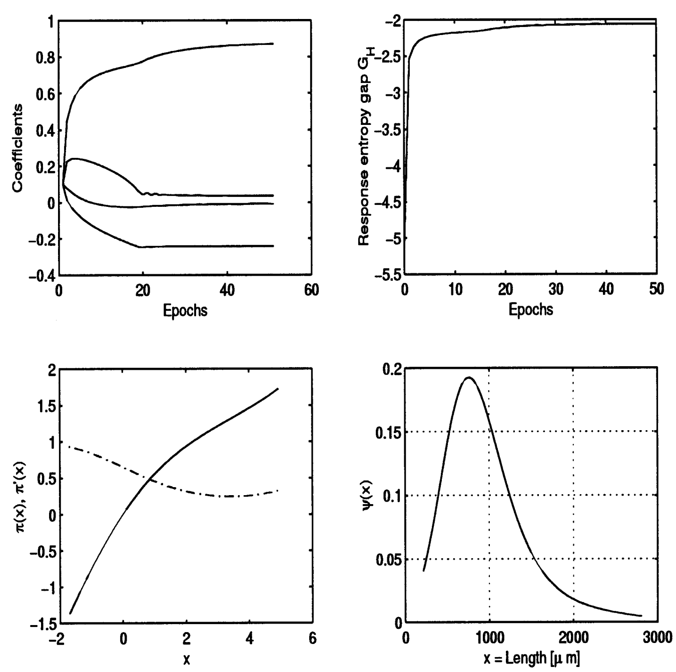

Fig. 5. PDF approximation of lengths of flax fibers embedded in a polypropylene slab for $\alpha = 0.60$. (Graphical symbols and conventions as in Fig. 4.).

learning phase was carried out for $E = 50$ epochs. The learning stepsize for any coefficient is equal to $\eta(t) = \eta_0 \exp(-1.5t/E)$, with $\eta_0 = 0.001$ and t being the epoch index.

With these learning conditions fixed, the first step is parameter α validation: The learning process has been repeated for every value of α in the range [0.1, 2.5] with step 0.05; in order to monitor the behavior of the algorithm, the ensemble-average estimation $\langle g_H(p, b) \rangle$ of the

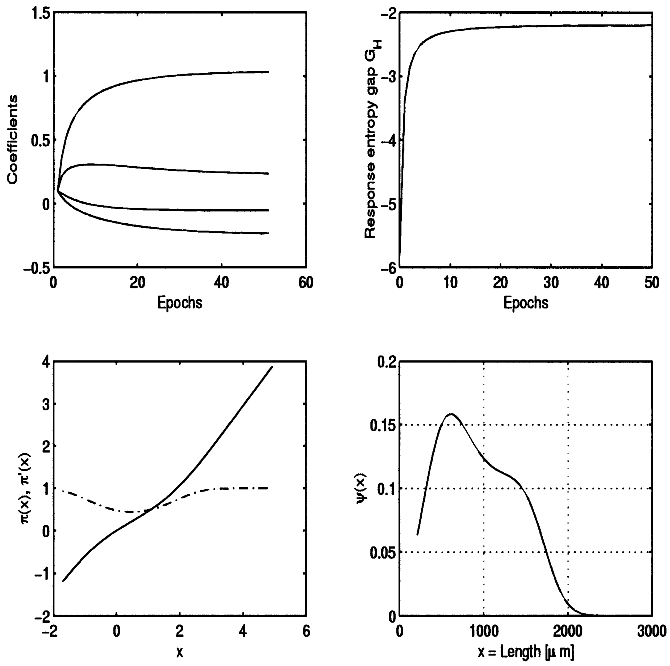

Fig. 6. PDF approximation of lengths of flax fibers embedded in a polypropylene slab for $\alpha = 0.90$. (Graphical symbols and conventions as in Fig. 4.).

entropy-gap G_H has been computed at the end of each 50-epoch-long trial and stored. The result of this preliminary experiment is shown in the Fig. 3. As long as $\alpha \in [0.25, 0.9]$ the entropy assumes the highest values, with a maximum in $\alpha = 0.35$.

After validation, the learning phase may be carried out again with some values of α belonging to the mentioned interval in order to monitor closely the behavior of the neuronal structure and of learning theory. In particular, it is worth monitoring the values of the $\bar{N} = 4$ polynomial's coefficients and of the ensemble-average estimation of the entropy gap at the end of every learning epoch, as well as to observe the shape of functions π, π' and ψ at the end of the whole learning process. These results are reported in the Figs. 4–6 corresponding to the values $\alpha = 0.35, 0.60$, and 0.90, respectively.

The experience leads to the conclusion that, among the examined density estimates, which result to be the most entropic ones, the estimate pertaining to value $\alpha = 0.6$ is the more reliable one.

V. Conclusion

A new artificial neuron model, endowed with adaptive activation function, as been presented, which tunes through an information-theoretic based learning paradigm. It is devoted to low-order, smooth, robust probability density function estimation and has been designed in a polymer-composite characterization context, that the presented experimental results pertain to.

We mention a problem regarding the selection of a suitable level of smoothness in the resulting probability density function estimate: A limitation of the method is that the validation analysis for parameter α produces a range of admissible values, thus, its selection may only rely on accumulated experience on the considered applied field.

The proposed PDF-estimation method is currently under further investigation in the context of polymer-composites mechanical features statistical modeling which involves both PDF-estimation and neural-network based statistical black-box identification of complex mechanical system.

Acknowledgment

The author wishes to thank Dr. J. Biagiotti for the kind assistance in the interpretation of the experimental results and former students C. Vedova, A. Cardoni, R. Donati, and E. Donati for sharing measures on polypropylene slabs.

References

[1] M. Battisti, P. Burrascano, and D. Pirollo, "Efficient minimization of the KL distance for the approximation of posterior conditional probabilities," *Neural Processing Lett.*, vol. 5, pp. 47–55, 1997.
[2] J. Biagiotti, S. Fiori, L. Torre, M. A. López-Manchado, and J. M. Kenny, "Mechanical properties of polypropylene matrix composites reinforced with natural fibers: A statistical approach," Polymer Composites, to be published.
[3] A. N. Birkett and R. A. Goubran, "Nonlinear adaptive filtering with FIR synapses and adaptive activation functions," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1997, pp. 3321–3324.
[4] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.
[5] P. W. Buchen and M. Kelly, "The maximum entropy distribution of an asset inferred from option prices," *J. Finance Quantitative Anal.*, vol. 31, pp. 143–159, 1996.
[6] P. Burrascano, E. Cardelli, A. Faba, S. Fiori, and A. Massinelli, "Application of probabilistic neural networks to eddy current non destructive test problems," in *Proc. 7th Int. Conf. Eng. Applicat. Neural Networks (EANN'2001)*, Cagliari, Italy, July 16–18, 2001, pp. 192–195.
[7] P. Coppens, *International Tables for Crystallography*, U. Shmueli, Ed. Dordrecht, The Netherlands: Kluwer, 1993, vol. B, Reciprocal Space, sec. 1.2.11.
[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[9] M. M. Dastani, "Functie-Benoderiungmet Feed-Forward Netwerkwen," Master's thesis, Univ. Amsterdam, 1991.
[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood for incomplete data via the EM algorithm," *J. Roy. Statist. Math. Soc.*, vol. B 39, no. 1, pp. 1–38, 1977.
[11] W. Duch and N. Jankowski, "Survey of neural transfer functions," *Neural Comput. Surveys*, vol. 2, pp. 163–212, 1999.
[12] S. Fiori and P. Bucciarelli, "Probability density estimation using adaptive activation function neurons," *Neural Processing Lett.*, vol. 13, no. 1, pp. 31–42, Feb. 2001.
[13] S. Fiori, "Probability density function learning by unsupervised neurons," *Int. J. Neural Syst.*, vol. 11, no. 5, pp. 399–417, Oct. 2001.
[14] ——, "Entropy optimization by the PFANN network: Application to independent component analysis," *Network: Comput. Neural Syst.*, vol. 10, no. 2, pp. 171–186, May 1999.
[15] ——, "Blind signal processing by the adaptive activation function neurons," *Neural Networks*, vol. 13, no. 6, pp. 597–611, Aug. 2000.
[16] ——, "Hybrid independent component analysis by adaptive LUT activation function neurons," *Neural Networks*, vol. 15, no. 1, pp. 85–94, Jan. 2002.
[17] ——, "A contribution to (Neuromorphic) blind deconvolution by flexible approximated Bayesian estimation," *Signal Processing*, vol. 81, no. 10, pp. 2131–2153, Sept. 2001.
[18] ——, "Information theoretic learning for FAN network applied to eterokurtic component analysis," *Proc. Inet. Elect. Eng. Vision, Image, Signal Processing*, vol. 149, no. 6, pp. 347–354, Dec. 2002.
[19] P. M. Mallick, *Fiber-Reinforced Composites*. New York: Marcel Dekker, 1993.
[20] V. K. Rohatgi, *Statistical Inference*. New York: J. Wiley, 1984.
[21] A. Silvescu, "Fourier neural networks," in *Proc. Int. Joint Conf. Neural Networks*, vol. 1, 1999, pp. 488–491.
[22] A. Sudjianto and M. H. Hassoun, "Nonlinear Hebbian rule: A statistical interpretation," in *Proc. Int. Conf. Neural Networks*, vol. 2, 1994, pp. 1247–1252.
[23] N. A. Vlassis, G. Papakonstantinou, and P. Tsanakas, "Mixture density estimation based on maximum likelihood and sequential test statistics," *Neural Processing Lett.*, vol. 9, pp. 63–76, Feb. 1999.
[24] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Networks*, vol. 3, pp. 889–898, Nov. 1992.